SFA Data Standards

I. Scope

All SFA data implementations at the Virtual Data Center are subject to the standards in this section.

II Applicability

These standards apply to all SFA data implementations under the Modernization Blueprint. This is not intended to be applied retroactively to legacy systems unless legacy modifications are being made to accommodate Modernization Blueprint objectives.

III Meta Data Standards

Data is an SFA asset just as important as its people, finances, equipment and intellectual assets. Indeed, it is hard to imagine any aspect of SFA operations that does not make heavy use of its current and historical information. Both the effectiveness and efficiency of the SFA organization are inextricably intertwined with its data resources. It is imperative that SFAs data asset is responsibly managed. A central component to responsible management of data is its meta data. The term "meta data" is used to mean descriptive information about data assets. This section contains the standards for SFA metadata.

III.1 Logical Data Requirements Defined

Logical data requirements are information needs by the business stated in terms independent of any technical considerations. Logical data requirements are represented in a logical data model (or its object-oriented equivalent) and are comprised of the following components:

- Subject Areas

  Definition:

  A subject area represents a primary group of information, which centers on a major resource, product or activity related to the management of the enterprise. Subject Areas are used as the basis for developing and analyzing logical data models. They are extremely useful when working with complex data models or when viewing data from a particular perspective. Subject areas are comprised of one or more non-overlapping entities. While subject areas may be hierarchically defined (subject areas defined as containing other subject areas) this practice is not supported in the current SFA environment. Subject areas may only be used in logical data models if the have been previously defined in the Conceptual Enterprise Data Model.

- Entities

  Definition:

  An entity is a fundamental thing of relevance to the enterprise about which data should be kept. An entity is the collection of entity occurrences that share a common definition, as well as common characteristics and relationships. An entity type may be a role played by a person (i.e., an EMPLOYEE, a CLIENT), a place, thing, resource, concept, or event. An entity type should always be placed within the boundaries of one and only one subject area.

- Attributes

  Definition:

  Attributes are the details relevant to the entity being described. An attribute that is maintained, as a predicate of an entity is a detail that specifically provides detail about an occurrence of that entity. An attribute can and should not have attributes of its own. When an attribute implies that it has attributes in itself it is called a *composite* or intelligent attribute. When this occurs it is because it has not been fully normalized and usually results in a loss of business information.

- Relationships

  Definition:

  A relationship is a relevant business association between occurrences of an entity and itself or another entity. A relationship between two entities is called a non-recursive relationship. An involuted (recursive) relationship is a relationship that describes the association between occurrences of the same entity type. An example of an involuted relationship in manufacturing is PART (made up of other parts). Every relationship must be named.

III.2 Meta Data Standards for Logical Data Requirements

This section details the information or meta data that is to be collected for logical data objects. Some of this information is typically collected during the Definition phase of the systems development life-cycle while other information is not known until the end of the Design stage of the Construction phase. The meta data provided by projects to the CIO/IT Management/Data Quality Management team will be recorded in the Conceptual Enterprise Data Model (CEDM).

The standard lists meta data "categories" such as Name, Description, Length, etc. Before an entity or attribute is added to the CEDM, certain categories of meta data must be provided according to the level of obligation on each meta data category as follows:

M – Mandatory - always required
C – Conditional - required to be present under certain specified conditions
O – Optional - allowed but not required
N/A – Not Applicable at this time

Currently, meta data standards exist for the following logical data objects:

- Subject Areas
- Entities
- Attributes
- Relationships

Subject Area Meta Data Standard

Metadata Category	CEDM Obligation	Description
Subject Area Name	M	The logical name for a new subject area . All candidate Subject Areas must be approved by CIO/Data Quality Management. A subject area must have at least one entity specified. Once approved the subject area will included in the CEDM and reflected in the Modernization Blueprint.
Definition	M	A business meaningful, narrative description of the scope of the information contained in the subject area. This description should unambiguously define the category of information (entities) scoped for the subject area. The subject area may not overlap the scope of information contained in any other CEDM subject areas.
Strategic Steward	O	The Strategic Steward, if known, as defined in Section II.3.1.2 Stewardship Roles and Responsibilities. It is assumed that all entities in the subject area have the same Strategic Steward unless exceptions are noted in the Strategic Steward specified for an individual entity.

ENTITY Meta Data Standard

Metadata Category	CEDM Obligation	Description
Entity Name	M	The logical name for the normalized logical data requirement
Access Name	C	This is not required during the Definition phase of the SDLC. It is required upon completion of the Construction phase of the SDLC. The physical name corresponding to the Entity Name (i.e., the database instance and table name under which it is physically implemented in the database, or the file name in which the data is actually stored.)
Definition	M	A business meaningful, narrative description of what the entity contains. This description indicates what the entity is, not how it is used.
CEDM Subject Area	M	The CEDM subject area under which the entity is categorized.
Strategic Steward	O	The Strategic Steward if known, as defined in Section II.3.1.2 Stewardship Roles and Responsibilities, for an entity. This is only specified when the Strategic Steward for the entity is different from the Strategic Steward for the CEDM Subject Area.

Metadata Category	CEDM Obligation	Description
Business Steward	O	The Business Steward, if known, as defined in Section II.3.1.2 Stewardship Roles and Responsibilities. It is assumed that all attributes in the entity have the same Business Steward unless exceptions are noted in the Business Steward specified for an individual attribute.
Technical Steward	O	The Technical Steward, if known, as defined in Section II.3.1.2 Stewardship Roles and Responsibilities. It is assumed that all attributes in the entity have the same Business Steward unless exceptions are noted in the Technical Steward specified for an individual attribute.
Comment	O	Additional narrative description of an entity that indicates the business rules, explanatory information on how the information is used, etc.
Modeler's Notes	C	Modeling notes include information relevant to data modeling with regard to technique or alternatives. Validation notes indicate specific uncertainties regarding the business rules reflected in the model. A subject matter expert validates the questions that arise from these notes. This field is eventually deleted from the CEDM meta data once the rules have been validated.

Attribute Meta Data Standard

Metadata Category	CEDM Obligation	Description
Attribute Name	M	The label of an attribute using business-meaningful language.
Access Name	C	This is not required during the Definition phase of the SDLC. It is required upon completion of the Construction phase of the SDLC. The physical table/column or file/field name corresponding to the attribute .
Definition	M	A business meaningful narrative describing the meaning of the attribute. It should describe what it is, not how it is used, derived, etc. Attributes should be described within the context of the business function being supported.
Business Steward	O	The Business Steward, if known, as defined in Section II.3.1.2 Stewardship Roles and Responsibilities. It is assumed that all attributes in the entity have the same Business Steward unless exceptions are noted in the Business Steward specified for an individual attribute.
Technical Steward	O	The Technical Steward, if known, as defined in Section II.3.1.2 Stewardship Roles and Responsibilities. It is assumed that all attributes in the entity have the same Business Steward unless exceptions are noted in the Technical Steward specified for an individual attribute.
Data Type	M	The way domain values are stored in a database (e.g., integer, short(small) integer, decimal, character, variable character, and date&time, date, binary).
Security Level	N/A (future)	A classification that shows the level of protection required to prevent disclosure/misuse. *CEDM classifications to be determined.*
Maximum Character Count	M	The field length of the attribute. For the CEDM, it should be large enough to accommodate all requirements, yet precise enough to allow for accuracy.
Update Frequency	O	A description of the frequency of updates to the domain (e.g., daily, weekly, monthly, quarterly, annual, as needed/ no definite pattern – usually used for code tables). This information informs implementers and/or database administrators when to refresh their data.
Update Source	C	The name(s) of the system(s) or organizations that currently creates or update this attribute. For legacy

Metadata Category	CEDM Obligation	Description
		systems, the source system(s) are listed. For new attributes: projects in the Definition Phase system designations may not yet have been determined, for Construction Phase the Update Source should be provided.
Derivation	M	Indicates whether the attribute is atomic (i.e., describes a single concept and is not derived using a formula) or, if not atomic, the category of derivation. There are three possible options for describing an attribute, column or data element: **Atomic:** attributes that are not derived (i.e., the result of computational operations) and do not contain multiple concepts. **Composite**: attributes that describe multiple concepts (e.g., a single attribute called "name" that is composed of the concatenation of first and last names). When identifying a composite attribute that is required to be used within a system, all pieces of data which make up this composite data element should be listed in the *Formula* metadata category (see below). **Derived**: attributes that represent the results of computational operations performed on other attributes (e.g., salary is derived from number of hours times hourly rate). The computations may involve algorithms supported by two or more attributes within a single entity instance, or algorithms summarizing attribute values across multiple entity instances within a single entity or across multiple entities. The algorithm should be listed in the *Formula* metadata category (see below).
Formula	C	A narrative description of the algorithm used to develop attributes that are derived or composite.
Values	C	The domain of all actual codes and their descriptions (e.g., for marital_status_code, M = Married, S = Single, D = Divorced). The values should be included in the code attribute as well as the code description attribute.
Edits	C	Contains the list of valid values or edit criteria used to

Metadata Category	CEDM Obligation	Description
		accept/reject data.
Authority Reference	C	The official regulation, policy, guidance, etc. that specifically requires the Department of Education to capture, maintain, exchange this data. The text in this reference must directly reference the data. For legacy systems, this information is provided if it is easily identified.
Comment	O	Additional narrative description of an attribute element. This includes the method of creating and maintaining IDENTIFIERs when proposed as primary key attributes.

Relationship Meta Data Standard

Metadata Category	CEDM Obliga-tion	Description
Relationship Name	**M**	A verb or verb phrase that describes the fundamental nature or business meaning for associating two entities. The relationship name should be stated in terms that are meaningful to the business.
Relationship Optionality	**M**	A staement of the optionality of an entity occurrence associated with the other entity in the relationship. Must be stated as 'mandatory' or 'optional'. Must be stated for each entity in the relationship.
Relationship Cardinality	**M**	The expected number of occurences of a related entity for a specific occurrence of a entity in a relationship. Must be stated as 'one' or 'many'. Must be stated for each entity in the relationship.
Comment	**O**	Additional narrative description of an attribute element. This includes the method of creating and maintaining IDENTIFIERs when proposed as primary key attributes.

IV. Logical Data Object Naming Standard

A data-naming standard is a set of rules that assists in the proper management of corporate data. Data names should be designed to minimize subjectivity and to promote common and consistent interpretation of their meanings. Benefits of a naming standard include:

- Universal identification of general business subject matter
- Quick identification of the class of business data
- Unambiguous and descriptive data names
- Easy resolution of potential synonyms, homonyms and aliases

The SFA logical naming standards contained in this section cover the following topics:

- Data Object Naming Components
- Basic Naming Rules
- Abbreviation Usage
- Data Name Format

IV.1 Data Object Naming Components

Logical data objects (subject areas, entities and attributes) are made up of three types of name components as follows:

Component	**Definition**
Modifier Word	A word that further defines and renders a data name unique when used with a Prime word and Class word. Usually this is an adjective but occasionally it may be a Prime word.
Prime Word	A word included in the name that represent the subject of the logical data object.
Class Word	A word included in the name, which classifies the kind of information being retained. Class words are only applicable to the attribute logical data object.

A data object name is formatted as follows:

Prime Word(underscore)Modifier Word(underscore)Class Word
One or more Modifier words may be used.
Example- Student_Application_Effective_Date

Provided below is a summary of the Naming Component requirement per Logical Data Object Type (m-mandatory, o-optional, n/a-not applicable):

Logical Data Object	***Prime Word***	***Modifier Word***	***Class Word***
Subject Area	M	N/A	N/A

Entity	M	O	N/A
Attribute	M	O	M

IV.2 Data Object Naming Rules

The basic rules for naming logical data objects are as follows:

- All names may not exceed a maximum of 32 characters in length. Prime words, Modifier words and Class words must be separated by an underscore (_)
- With the exception of the underscore, special characters are not allowed
- All names must start with a Prime word.
- Only attribute names use Class words. An attribute name must end with an approved Class word as defined in this standard.
- Prime and Modifier words should enhance the readability of the object to a non-technical business user
- Associative entity names should include the prime word from each parent entity unless a more descriptive business meaning could be conveyed using another word
- Nouns must be in a singular form

IV.3 Abbreviation Usage

Abbreviations should only be used if the name you are constructing does not fit within the standard name length constraint documented in the basic rules stated above. The order of abbreviations used should be from the last word to the Prime word. Each word is abbreviated until the name meets the 32 character length.

IV.4 Data Naming Format

IV.4.1 Prime Words

Prime words are nouns used in industry or business that are common and standard for that particular business channel or industry. They are words that one can normally hear during a typical business conversation within the organization.

Prime Word Examples:

Student	Institution	Guarantor
Loan	Regulation	School
Transcript	Curriculum	Lender

IV.4.2 Modifier Words

Modifier words are common adjectives used to quantify, clarify, convey distinctions or otherwise differentiate Prime Word data names. Although these words are usually adjectives some Prime words can also function as a modifier.

Repaid	Public	Active
Awarded	Governmental	Financial

The following table shows examples of Prime words that are subject nouns and used as Modifier words.

Student Loan	School Curriculum	Student Transcript
School Loan	Guarantor Loan	Lender Institution

IV.4.3 Class Words

Class words are only applicable to attribute type data objects. The purpose of the Class word is to group like types of data into a limited number of sets to standardize terminology and enhance readability. The Class word identifies and describes the general purpose or type of data object being named. They are self-explanatory; for example, the purpose of the Class word "identifier" indicates that the purpose of the attribute is to serve as the unique identifier for some other data.

The rules for Class words are as follows:

- Each attribute must end in a class word.
- Only one class word can be used in each attribute name.
- IDENTIFIER takes precedence over other class words – it is a non-intelligence bearing attribute whose purpose is to uniquely identify an entity in an entity type[1].
- If there are two or more class words that could be used to describe the attribute, the most encompassing one should be selected.
- Class words are reserved words. Entities or Relationships should not use a class word in its name.

Class Word	Description
Amount	A monetary quantity. Always expressed in whole and fractional portions.
Category	A classification of data that is not codified and does not require a reference or translation table to become meaningful information
Code	A differential of a classification of data that requires a reference or translation table to become meaningful information
Count	An integer number indicating a measure in the indicated unit of measure and available for arithmetic use

[1] For example, assume an attribute exists in a legacy system entitled STUDENT-NUMBER and there is no business significance to the composition of that attribute other than to uniquely identify a STUDENT. If it is selected as the identifier for STUDENT, both number and identifier are class words. Since only one class word can be used per attribute, the attribute name would become STUDENT-IDENTIFIER since identifier takes precedence over number as a class word.

Class Word	
Date	A measurement of time from which year, month and day may be determined
Dimension	A quantitative measure of spatial proportions in up to three dimensions
Identifier	A non-intelligence bearing attribute whose purpose is to uniquely identify an entity. This classword is only to be used for attributes making up an entity primary key and where no enterprise concensus has been achieved. Once an enterprise concensus for an entity identifier has been specified by the Business Steward for the entity then this classowrd must be changed to one of the other classwords in this list.
Image	A picture or graphic
Indicator	A simple Boolean flag (for example, Y = yes and N = no)
Name	*A word or phrase that constitutes a distinctive designation for a person, place, thing, concept or event*
Number	A numeric integer used for identification or sequencing and not intended for arithmetic use
Percent	A unit-less measurement expressing a part to a whole
Quantity	A real number indicating a measure implied to be in units and available for arithmetic use
Rate	A measurement of change over time expressed in designated units of measure
Text	Any of the forms, versions or editions in which a written work exists. Data having undefined, free-form, unstructured or unformatted content, including text and other printable characters
Time	An indication of time of day which is capable of indicating hours, minutes and/or seconds, including fractions

IV.4.4. Naming Standards Considerations for Data Model Components

In addition to the standards set forth above, model component specific standards exist for the following data model components of a Logical Data model:

Subject Areas
Entities
Attributes
Relationships

Subject Areas
The subject area name must be a singular noun, which is descriptive of primary business concept encapsulating the class of entities it includes. The name must be stated in business terms independent of any technical or automation considerations.

Entities
An entity name must be a singular noun or noun phrase in the singular form. These should be common nouns, not proper nouns. The word 'type' should not be included in it's naming. In a logical data model they should be in a phrase that clearly and uniquely identifies the information being described. It must represent a business concept independent of any technical or automation considerations. In a physical data model technical terms may be used if a business term does not satisfy the need.

Attributes
An attribute should be named as a noun or noun phrase and describe the purpose or content of that attribute. The name should be in recognized business terms. Generally speaking an attribute name should not reflect an order of occurrence (i.e., Address 1, Address 2, etc.) as it reflects an un-normalized view of that entity.

Relationships
The name of a relationship membership describes the reason for joining two entities of the types participating in the relationship. The name should be formed as a verb (either active or passive) that connects one entity type (the subject) to another (the object). The membership name forms the basis of a stated business rule. The relational pairing, since it can be read from either end, requires two membership names.